Haab, Timothy C., Interis, Matthew G., Petrolia, Daniel R., Whitehead, John C. (2013). From Hopeless to Curious? Thoughts on Hausman's 'Dubious to Hopeless' Critique of Contingent Valuation, *Applied Economic Perspectives and Policy 35*(4):593-612. Publisher version available from http://www.oxfordjournals.org/en/ (ISSN: 2040-5804) DOI: 10.1093/aepp/ppt029

# From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation

Timothy C. Haab*, Matthew G. Interis, Daniel R. Petrolia, and **John C. Whitehead**

**Abstract**

Hausman "selectively" reviewed the contingent valuation method (CVM) literature in 2012 and failed to find progress in the method during the 18 years since Diamond and Hausman argued that unquantified benefits and costs are preferred to those quantified by CVM. In this manuscript, we provide counterarguments to Hausman's claims, not with the intent to convince the reader that the debate over CVM is settled in favor of the method, but rather to argue that the intellectual debate over CVM is ongoing, that dismissing CVM is unwarranted, and that plenty of work remains to be done for the truly curious researcher.

**Key words:** Contingent valuation method, Nonmarket valuation

The contingent valuation method (CVM) has been under scrutiny ever since the high profile of the role of contingent valuation in litigation following the Exxon Valdez oil spill (Carson 2011). Contingent valuation was developed because revealed preference methods (e.g., travel cost) that had been used to measure non-market values were limited. Benefit-cost analysis, and other types of policy analysis, required empirical estimates of value that were still not measureable. Contingent valuation employs surveys containing detailed descriptions of a hypothetical valuation scenario, payment vehicles (e.g., taxes), payment rules (e.g., majority rule) and valuation questions (e.g., referenda). The initial motivation for the method came from U.S. federal government agencies from the 1940s to the 1980s; the National Park Service sought estimates of the economic value of park recreation and the Army Corps of Engineers was seeking estimates of water-based recreation (to justify dam building), while the U.S. Environmental Protection Agency sought empirical estimates of the value of risk reduction in environmental and health economics.

The major development that launched the current controversy was the recognition that passive-use values, described by Krutilla (1967), could be estimated using contingent valuation. Existing revealed preference methods (e.g., travel cost) could not capture passive-use value, a potentially key component of total non-market value. Thus, stated preference methods, including CVM, remain the only known set of methods for capturing passive-use values.

The mid 1970s through the 1990s was a period of intense methodological development and increased acceptance of CVM, and saw a broadening of applications beyond recreation and health, an expansion to developing countries, and the publication of a number of influential overviews. First among these was the article by Randall et al. (1974) in the first issue of the Journal of Environmental Economics and Management, which introduced more differentiation of contingent valuation surveys from opinion polls by using photographs to help describe the valuation scenario.

Two key milestones in the broader acceptance of CVM were the publication of the Mitchell and Carson (1989) book, which first integrated economic theory, survey research methods, and social science measurement issues, and the Exxon Valdez oil spill. The Exxon Valdez oil spill first brought contingent valuation to the attention of many economists, government agencies and the courts. In 1993, the U.S. National Oceanic and Atmospheric Administration (NOAA) assembled a “blue ribbon panel” to assess the method. The panel concluded that, under the right conditions, the method could play a legitimate role in damage assessment, including the estimation of lost passive-use values (Arrow et al. 1993). The panel’s recommendations created a number of testable hypotheses that emerged in the literature of the 1990s and 2000s. In response to the NOAA Panel’s conditional endorsement of the contingent valuation method, a group of primarily Exxon-funded researchers published a compendium of critiques questioning the accuracy and validity of the method (Hausman 1993).

Following this controversy, the contingent valuation method was the subject of a 1994 symposium in the American Economic Association’s Journal of Economic Perspectives. Portney (1994) introduced the important issues, Hanemann (1994) defended the method, and Diamond and Hausman (1994) argued that, for the purposes of damage assessment and government decision-making, unquantified benefits and

costs are preferred to benefits and costs quantified by the CVM, and that expert opinion is preferred to imperfect contingent valuation estimates of value. While Diamond and Hausman raised a number of important issues, their negative opinion has done little to quell the demand for contingent valuation research.

In the aftermath of the BP Deepwater Horizon oil spill in 2010, the Journal of Economic Perspectives held a symposium that revisited the nightmare of the "CVM debate" that followed the Exxon Valdez oil spill: Kling, Phaneuf, and Zhao (2012) provide a balanced interpretation of the usefulness of the CVM; Carson (2012) is optimistic about the progress researchers have made in the past twenty years, which suggests that under the right circumstances the CVM can be a useful valuation tool; Hausman (2012) also "selectively" reviews the CVM literature, and fails to find progress during the eighteen years since Diamond and Hausman (1994). Hausman goes on to outline a series of potentially fatal flaws in the method, and ultimately declares the CVM "hopeless."

To be sure, Kling, Phaneuf, and Zhao (2012) and Carson (2012) offer some counterarguments to the assertions made by Hausman (2012), but given the coincidental nature of the three papers, they do not provide a specific critical response to the potentially damning critique by Hausman. Indeed, Hausman acknowledges his selective reading and interpretation of a vast literature that has developed in the last twenty years.2 The curious reader is directed to the comprehensive bibliography of over 7,500 scholarly papers and studies on contingent valuation provided by Carson (2011). The vast majority of the reviewed papers were written during the 18 years since the initial Journal of Economic Perspectives symposium on Contingent Valuation in 1994, where Hausman first pronounced upon the method. Yet Hausman's (2012) claim that "three long-standing problems continue to exist," namely, hypothetical bias, the divergence between willingness to pay (WTP) and willingness to accept (WTA), and the lack of scope effects, requires a more direct and thorough response. We do so not with the intent to convince the reader that the debate over contingent valuation is settled in favor of the method— we ourselves remain cautious on many fronts—but rather to emphasize to non-specialist economists and others that the intellectual debate over contingent valuation is far from over, and that there remains plenty of research to be done by the intellectually curious. This is something we would hope even a skeptic would embrace.

In short, we disagree with, and more importantly provide evidence that contradicts, Hausman's (2012) conclusion that "despite all the positive-sounding talk about how great progress has been made in contingent valuation methods, recent studies by top experts continue to fail basic tests of plausibility." Instead, we find hope in recent studies by respected experts that contingent valuation can indeed provide plausible value estimates as well as information on market extent and public opinion, and as a consequence should be one (but not the only) tool at the disposal of policy makers and litigants alike. This remains especially true in contexts where important components of the benefits of a proposed public policy or the damage claims in a lawsuit lie beyond the reach of evidence from existing markets. It is sometimes much harder to measure benefits and damages in these contexts, but this does not diminish the imperative that we do so if economic criteria are to be honored in social decision-making. Further, we find that critical examination of many of the papers that Hausman

uses as evidence against CVM in fact raise important questions about any type of value elicitation survey, real or hypothetical, and in some cases provide support for hypothetical surveys when compared to real monetary exchanges. In short, we find promise for the curious researcher that the CVM debate is not settled, important questions remain, and that a critical examination of the CVM literature will provide fertile ground for future research.

## Three Long-Standing Problems

*Hypothetical Bias*

Hausman (2012) examines three main issues that he believes continue to plague stated preference studies. The first issue he discusses is hypothetical bias, where "what people say is different from what they do." Most hypothetical bias studies, whether they take place in environmental valuation, marketing, or public polling, use some form of the stylized testable hypothesis that stated preference responses are identical to responses to analogous questions when money or some other real outcome is at stake. If the hypothesis is rejected, the stated preference study is viewed as flawed. While we respond more specifically below to Hausman's specific criticisms, we note that despite issues of dealing with the very hypothetical nature of intended behavior questions, hypothetical questions are regularly used to evaluate real world decisions outside of the realm of environmental valuation. Marketers frequently utilize hypothetical market situations to assess the demand for new products, the expansion of existing markets beyond traditional demographics, or to assess price changes beyond the realm of market experience (Morwitz, Steckel, and Gupta 2007). Pollsters regularly rely on hypothetical questions of intended voting behavior to predict election outcomes and understand voter preferences for campaigns (Hillygus 2011). Despite the widespread use of hypothetical questions by private parties with real stakes and the commitment of real resources on the outcomes of such surveys, discussions of hypothetical bias draw special ire when placed in the context of benefit-cost analysis and the potential for use in litigation and public policy decisions (Portney 1994).

Hypothetical bias has been examined in numerous empirical studies, and several meta-analyses compare value estimates from hypothetical and real choices. For example, List and Gallet (2001) find that estimates based on hypothetical choices are on average about 3 times higher than those based on real choices, but that the discrepancy depends on the elicitation mechanism, the value elicited (WTP or WTA), and whether the good is public or private. Little and Berrens (2004) find that hypothetical estimates are about 3.13 times higher than those based on real choices, and find that referenda and correcting for respondent uncertainty can reduce this discrepancy. Murphy et al. (2005) find hypothetical values are on average 1.35 times higher than those based on real choices. While these studies control for the elicitation mechanism, they do not look more specifically at respondent incentives, so it is unclear whether the discrepancies found should be expected based on economic theory.

Two recent papers have fundamentally changed researchers' thinking about hypothetical bias. Carson and Groves (2007, 2011) show that without a closer examination of the incentive structure of a particular stated preference survey, we cannot actually predict whether distortions due to hypothetical bias should or should not

occur (nor the magnitude of any predicted bias), because economic theory of how people make choices relies on knowing the incentives faced by the respondents. Placed primarily in a theoretical context, Carson and Groves' arguments regarding incentive properties open a new exciting line of research for applied, behavioral, and experimental researchers to investigate the degree to which the incentive properties of various question formats can reduce or increase hypothetical bias.

Hausman cites three studies which find that "intentions to purchase new products" are overstated: Jamieson and Bass (1989), Hsiao et al. (2002), and Morwitz et al. (2007).As noted by Carson and Groves (2007),we should typically expect that stated intentions to purchase new products will overstate true intentions. The logic is that by choosing "yes" (I would buy this product), the respondent increases the likelihood of having the product available for possible future purchase, under the reasonable assumption that the survey is being used to determine whether the new product should be brought to market. Utility is derived from an expansion of the individual's future choice set. The Jamieson and Bass (1989) study and the Hsiao et al. (2002) study involve products that are already available, but are relatively new to the market.3 In these cases, it is unclear what exactly the incentives are for the respondent, since responses do not influence whether the product is brought to the market and thereby do not expand the available set of choices. As Morwitz et al. (2007) indicate, purchase intention studies concerning existing products are often used by manufacturers and suppliers when they need to decide about "whether to increase or reduce production levels, whether to change the size of the sales force, and whether to initiate a price change."4 However, it is not clear in this case exactly how responses will affect something the respondent cares about (i.e., product price). Thus, we should not be surprised by the findings of these studies (and others cited therein) that stated purchase intentions for existing products have some predictive ability, but that it is not always particularly strong.5 The issue is further muddled when moving from market goods to nonmarket goods, as is the case with most environmental applications of CVM surveys. It is perhaps surprising that Hausman did not more carefully consider the incentive structure of the surveys used in these cited studies because he points out that one possible cause of hypothetical bias is that the survey does not explicitly state how the survey responses will be used. In this case, respondent incentives are unknown, so there is no reason not to expect (or to expect) hypothetical bias.

While Hausman's use of these papers to provide evidence that intentions often overstate behavior is partially correct (e.g., Morwitz et al. [2007] examine the correlation between purchase intentions and actual sales of market goods, but they do not consider the magnitude of the quantity purchased), he misrepresents the purpose and conclusions of each. Morwitz et al. (2007) use meta-analysis to determine the conditions under which the correlation between purchase intentions and actual sales increases, and find that the correlation is higher for existing products, for durable goods, and when the time between the hypothetical scenario and real outcome is shorter. These authors conclude that "the results indicate that purchase intentions are predictive of future behavior, and that much of the variation in the intent-behavior relationship can be explained by the characteristics of the study." Hsiao et al. (2002) develop four econometric models and compares which does best at linking intentions with behavior. He concludes that there is "a remarkably stable relationship between intentions and

purchase over time which indicates that intentions are powerful predictor [sic] of actual purchase." Similarly, Jamieson and Bass (1989) attempt to move the literature forward by developing models that improve forecast accuracy. The spirit of each of these papers is constructive and inquisitive and is inappropriately used in a destructive way by Hausman. These efforts in the marketing literature parallel the efforts of stated preference researchers who have made strides in understanding and reducing the magnitude of hypothetical bias in a number of ways (Loomis 2011).

In addition to these marketing studies, we can cite at least two examples from the environmental economics literature where stated preference estimates of prospective demand under unexperienced future conditions have been shown to predict with some accuracy revealed behavior when the future conditions actually materialize. For example, Grijalva et al. (2002) conducted a predictive validity test of rock climbing trip behavior. Respondents were surveyed about their current revealed preference trip behavior and their stated preference behavior under future access conditions. Following the realization of the hypothetical scenarios, respondents were surveyed again. With the hypothetical closure of rock climbing areas, stated preference rock climbing trips fell. When the areas are actually closed, actual trips differed in the expected direction and by similar magnitudes. Similarly, Whitehead (2005) conducted a predictive validity test of hurricane evacuation behavior. Respondents were surveyed about their revealed preference evacuation behavior after low-intensity storms and about their stated preference behavior after both hypothetical low-intensity and hypothetical high-intensity storms. Two hurricanes followed the survey, and respondents were surveyed again to determine their actual behavior. Models using revealed and stated preference evacuation data forecast actual behavior with prediction errors between .015 and .15 on a zero-to-one probability scale.

One of the four key areas of future stated preference research noted by Kling et al. (2012) is the need for validity tests on consequential surveys. Following Carson and Groves (2007), a consequential survey is one in which respondents believe their responses will affect something that they care about. Only for consequential surveys with well-defined incentives can economic theory make predictions about how respondents should answer, and therefore whether one should expect hypothetical bias. In support of the efficacy of consequentiality in hypothetical value elicitation, Kling et al. (2012) cite Vossler and Evans (2009) and Landry and List (2007), who find no hypothetical bias when responses are consequential, and Poe and Vossler (2011),who conduct four consequential laboratory experiments, and also demonstrate a lack of hypothetical bias. The consequentiality paradigm is further supported by Vossler, Doyon, and Rondeau (2012) in a field experiment context, and Bulte et al. (2005), Herriges et al. (2010), and Vossler and Watson (2013) in a field survey context. Johnston (2006) compares voting behavior in an actual referendum to that in a hypothetical CV referendum on the same issue, and where the incentives are designed to be as similar as possible; he finds no evidence of hypothetical bias, and points out that most studies comparing hypothetical and real referenda do not have the same incentive structure.

In the absence of true consequentiality, a number of approaches to mitigating hypothetical bias have become popular (Loomis 2011). Some researchers find that the divergence between hypothetical and actual WTP is mitigated or eliminated through

ancillary survey mechanisms such as “cheap talk” survey design, or more recently, oaths. Cheap talk in the valuation context involves providing respondents with additional instructions that explicitly encourage them to treat the hypothetical scenario as if an actual monetary transaction were taking place (Cummings and Taylor 1999; List 2001).

More recently, researchers have investigated the efficacy of other ancillary survey mechanisms for reducing hypothetical bias. Jacquemet et al. (2013) use a variation of a “solemn oath” to tell the truth and find it effective in reducing the gap between real and hypothetical responses.6 In a CVM setting, Carlsson et al. (2013) find that an oath script has significant effects on hypothetical response behavior across individuals and across multiple countries that is consistent with what would be expected from reduced hypothetical bias. However, because their survey is a non-experimental CVM application, they provide no direct evidence that hypothetical bias is reduced relative to a real treatment. Jacquemet et al. (2013) apply their version of the oath in an induced value second-price auction experiment for dolphin protection, and conclude that “For induced value, we find the oath-only treatment induced sincere bidding behavior on average within a second-price auction; the other hypothetical and real treatments did not,” [italics added].

Still more researchers advocate for various forms of “scenario adjustment”, including controlling for and netting out respondent departures from the conditions described in the offered choice scenarios (Cameron and DeShazo 2013) or taking into account the level of certainty respondents have that they would actually make the same decision if the choice was real (Champ and Bishop 2001; Blumenschein et al. 2008). Adjusting for certainty can produce “hypothetical” results that are often close to “real” values, and these alternative levels of WTP allow for sensitivity testing when CVM estimates are used in benefit-cost analysis (BCA). The CVM literature has not reached a conclusion about which hypothetical bias adjustment approaches should be used.7 This is one of the areas that require more research. However, the current practice seems promising, which leads us to assert that a range of likely numbers (e.g., WTP estimates) that likely includes the most accurate values is certainly better than no number.

The Carson and Groves paradigm for assessing the predictive ability of stated preference surveys has been extremely influential.8 The concept of consequentiality is relatively new, and although Carson and Groves have established a clear definition of what constitutes a consequential survey, the empirical methods used to measure and control for it vary in the literature, with mixed results. More studies that compare hypothetical and real choices, but which also focus on incentive structure of the questions, are therefore needed.

*Willingness to Pay and Willingness to Accept*

Since the onset of stated value elicitation surveys, researchers have been troubled by the seemingly simple finding, perhaps obvious to some, that compensation demanded (WTA) regularly exceeds compensation paid (WTP). Based on this commonWTA/WTP gap, researchers have split along two lines in their conclusions. Some dismiss contingent valuation based on the naı̈ve premise that the gap represents a violation of basic economic theory and thus invalidates value elicitation surveys. Others remain

curious about the cause of the disparity and wonder whether the gap can be explained, either within the neoclassical framework, or by necessary extensions to this framework that draw upon psychological insights from behavioral economics. In dismissing contingent valuation, Hausman (2012) claims that "Basic economic theory suggests that [questions phrased to elicit willingness to pay to avoid a negative outcome and questions phrased to elicit willingness to accept the negative outcome] should give (approximately) the same answer . . ." Indeed, using income elasticity estimates from over 200 published studies comparing WTP and WTA, McConnell and Horowitz (2003) "conclude that the ratio WTA/WTP is too high to be consistent with neoclassical preferences." The McConnell and Horowitz results rely on existing studies, most (if not all) of which rely on the parametric specification of elasticities, which derive either implicitly or explicitly from assumed consistent neoclassical preference structures between WTP and WTA.9 McConnell and Horowitz (2003) remain open to the possibility that neoclassical preferences may not always be an appropriate assumption, noting that their results "should help in developing explanations and alternative models, since income effects are such a prominent part of economic models such as choice under uncertainty." In other words, the WTP/WTA disparity is not a death knell for CVM, but rather a call for the curious researcher to more closely examine the assumptions and structures leading to these seemingly anomalous results. As noted in the previous section, recent research into the incentive properties of question formats can influence WTP measures. Most existing studies of the WTP/WTA gap draw on question formats with potentially incentive-incompatible question formats or preference structures. As such, the impact of incentive properties on the WTP/WTA gap remains largely unexamined.

While Hausman's claim may be true in terms of the "basic economic theory" that prevailed prior to the discovery of the WTP/WTA gap, he ignores an expanding body of literature that includes contributions from those who have been curious about whether there is an explanation that provides both an updated theoretical framework and experimental evidence that the gap can be explained within a slightly more general version of basic economic theory—although the debate remains unsettled.

One illustrative strand of this literature (but by no means the only set of contributions) begins with Hanemann's (1991) *American Economic Review* piece in which he lays out a neoclassical explanation for the WTP-WTA gap. Hanemann notes that the difference between WTP and WTA, or more precisely the difference between Hicksian compensating and equivalent variations, hinges on the ratio of the income elasticity to the Allen-Uzawa elasticity of substitution. Hanemann goes on to argue that it is likely that the elasticity of substitution for unique goods, such as those often valued in contingent valuation surveys, is likely to be particularly small, thus increasing the expected difference between neoclassical compensating and equivalent variation. Shogren et al. (1994) use a series of market and nonmarket experiments to test Hanemann's neoclassical substitution hypothesis and conclude: "Our experimental results support [Hanemann's] argument that the degree of substitutability between goods may drive the difference between WTA and WTP measures of value." Revisiting the experiments of Shogren et al. (among others), Morrison (1998) finds that "even when allowing respondents to learn through repeated trials, controlling as much as possible for Hanemann's substitutability argument . . . the result of WTA exceeding

WTP not only remains, but remains strongly significant." Shogren and Hayes (1997) counter "that the WTP-WTA disparity depends on the auction institution, not on a deviation from neoclassical rationality as suggested by the endowment effect," again highlighting the importance of understanding incentives. Hausman's 2012 selective oversight of such work allows him to dismiss contingent valuation on the grounds of violations of basic economic theory noting that:

> Various efforts have been made to extend the neoclassical framework in a way that rationalizes the gap. Proponents of contingent valuation have attempted rationalizations of these differences, but have not overcome the findings of Diamond and Hausman (1994) or the results of Milgrom (1993). Both papers demonstrate that the attempts to rationalize the well-recognized and persistent disparity between willingness to pay and willingness to accept fail as a matter of economic theory and observed empirical outcomes.

In a recent exploration of the impacts of reference dependence (i.e., endowment effects), Knetsch (2010) notes, in a particularly relevant footnote: "One somewhat unfortunate consequence of the many reports of large disparities between WTA and WTP valuations in contingent valuation surveys, is that many, especially environmental economists, have taken these disparity findings to be the result of a problem of the inadequacy of contingent valuation methods to accurately measure people's values rather than taking it more seriously as one of people valuing losses more than gains." Knetsch goes on to argue that the appropriate measure of value (WTP or WTA) will depend on the initial assignment of property rights, perceived or otherwise. This creates potential difficulties in the interpretation of benefit-cost analysis. 11 However, the prevalence of a WTP-WTA gap, or the existence of endowment effects, is not a fatal flaw in the contingent valuation method in and of itself, nor is it unique to CVM. Knetsch (2010) cites several lab and natural experiment studies that also find a large gap.

Based on our own reading of the literature, we are not ready to close the WTP-WTA debate. And apparently, we are not alone:

> A subtle controversy exists in the literature. At issue is the existence and interpretation of a possible gap between willingness to pay ("WTP") and willingness to accept ("WTA"). . . . In spite of the enthusiastic interpretations of the WTP-WTA gap as a fundamental feature of human preferences . . . in fact there is no consensus about whether the literature, considered in its entirety, supports such interpretations (Plott and Zeiler 2005).

*Scope Test*

Hausman argues that contingent valuation studies tend to fail the "scope test," and those that do pass the test fail to pass it "adequately." At the same time, he acknowledges that "We do not know how large scope effects should be." We share Hausman's (2012) ignorance about what would constitute scope effect "adequacy." A simple theoretical model of WTP, a difference in expenditure functions with changes in quality or quantity, can be used to show that WTP is nondecreasing in quality or quantity (Whitehead, Haab, and Huang 1998). The size of scope effects may be limited by diminishing marginal utility or substitution among nonmarket goods or between nonmarket and market goods. For example, Rollins and Lyke (1998) find that people do

distinguish between existence goods of different scope, but due to diminishing marginal valuations for larger scope goods, estimated differences between the values of larger scope goods may be negligible unless sample sizes are adjusted accordingly. Additionally, Amiran and Hagen (2010) develop a formal model of scope with bounded utility functions and show that relatively small scope effects are not inconsistent with economic theory. The size of the scope effect in this model depends explicitly on the substitutability between market and nonmarket goods.

Hausman considers the "Diamond-Hausman adding up test," developed by Diamond (1996) to be the definitive scope test. To illustrate the adding up test and other scope tests, suppose there are three samples of CVM respondents, as depicted in figure 1.

Sample 1 is presented a choice or WTP question for A, and then a second question for A+B. Sample 2 is presented a choice or WTP question for B, given that A is already provided as part of the consumption bundle and payment has already been extracted (labeled as B | A). Sample 3 is presented with a single choice or WTP question for the joint provision of A+B. Success in an adding up test occurs when the sum of WTP from the first question posed to sample 1 and sample 2 is equal to the WTP of sample 3. The standard scope test in the CVM literature, known as the external (i.e., split-sample) scope test, is a comparison of values across the first question for samples 1 and 3. Willingness to pay for A+B should be greater or equal

**Figure 1** Samples used in various scope tests

| | Sample 1 | Sample 2 | Sample 3 |
|---|---|---|---|
| First WTP Question | A | B \| A | A + B |
| Second WTP Question | A + B | | |

to WTP for A. The so-called internal scope test is a comparison of the first and second WTP values for sample 1. Again, WTP for A+B should be greater than or equal to WTP for A. The adding up test imposes a specific structure on the preference function, which may not be appropriate. External and internal tests do not impose any such structure.

Hausman (2012) highlights the evidence provided by Desvousges, Mathews, and Train (2012), who review 109 studies that report external and internal scope tests, and find that 36% pass the test, 15% fail the test, and the rest have mixed results. Desvousges, Mathews, and Train (2012) find only a few studies that provide enough information to conduct an adding up test. The axiom of revealed preference (on the part of CVM researchers) suggests that external and internal scope tests are preferred to the adding up test. One reason for the rejection of adding up as an appropriate test by CVM researchers is the unnecessary structure imposed on preferences. Another reason is that the adding up test asks respondents in sample 2 to assume that A has been provided by the policy process and payment has already been extracted. This more complicated counterfactual scenario substantially increases the cognitive difficulty of the

valuation task (Hanemann 1994). Also, the adding up test requires three subsamples, which increases the cost of a study substantially. It is apparently not clear to many researchers whether the benefit of conducting the adding up test, relative to the more straightforward external and internal scope tests, is worth the cost.

Given the absence of any explicit adding up test for most of the 109 scope-effect papers in the literature, and the fact that the adding up test is not the only valid test for adequate sensitivity to scope, it is necessary to more closely examine the types of scope tests that are routinely conducted. When considering whether an individual study adequately passes a scope test, any meta-analysis should consider characteristics of the study that might drive scope-testing results. Smith and Osborne (1996) conduct a meta-analysis of studies focused on the value of changes in visibility at U.S. parks and find clear responsiveness of CVM estimates to scope. These authors warn, however, that scope predictions are sensitive to theoretical assumptions imposed during model estimation. Richardson and Loomis (2009) find that the size of the population change is statistically and economically significant in a meta-analysis of 67 WTP estimates of protection for threatened and endangered species. In two specifications, the scope elasticity of WTP is not statistically different from one. Ojea and Loureiro (2011) undertake a meta-analysis of 355 biodiversity contingent valuation studies that focus on the issue of scope-test satisfaction, and find that scope effects are more likely to be found when scope is measured absolutely instead of relatively.

The type of valuation question might play a role in scope sensitivity. Without providing empirical evidence, Carson and Groves (2007) argue that a consequential referendum is more likely to produce scope effects. This remains an open question for research. Empirical tests of the scope elasticity theory developed by Amiran and Hagen (2010) would also be useful. A meta-analysis based on the 109 studies reviewed by Desvousges, Mathews, and Train (2012), with further examination of what is meant by "mixed" results, is needed to determine the factors that influence scope failure.

### *Other Issues*

In addition to his three primary critiques, Hausman (2012) raises a number of other issues which we feel deserve further reflection.

### *The Use of Experts*

In response to questions about how to handle nonuse values in decision-making without recourse to stated preference surveys, Hausman advocates the use of experts: "[P]ublic policy will do better if expert opinion is used to evaluate specific projects, including non-use value . . ." This argument is also made in the original Diamond and Hausman (1994) paper. Beyond the arrogance of assuming that experts know better what the value should be than the public itself, the use of experts in the valuation process raises a number of new and important questions.

What kinds of experts should be included on the suggested expert panels? It is hard to imagine that non-economists would be better equipped than economists to assess passive use value, or even to be familiar with the concept in order to consider it in the first place. For ecological services, for example, are we to substitute ecologists'

subjective opinions about “intrinsic value” for the instrumental values required for benefit-cost analysis? How are these intrinsic values to be monetized? Why would experts be better than non-experts at determining the monetized nonuse values accruing to society? After all, decision-makers are usually interested in the preferences of the general public, so it is difficult to see howexperts, as a highly selective sample from the general population, should be better informed about social values than nonexperts, that is, a representative sample from the general public stating their own preferences directly. If we are discussing whether a given environmental project does or does not deliver a specific increment of physical ecological function, then generally an expert is better informed than the average person. But value is distinct from function. We might expect that the average person knows his own preferences better than the expert, conditional on full information. Whether we can accurately elicit that value from the public brings us full circle back to Hausman’s critique of methods addressed previously.

The more practical question, and really the most fundamental one, is how these alleged experts can determine nonuse values without having first conducted at least an implicit stated preference study as the basis for their expertise in judging societal nonuse values? It is well-known that the distinctive characteristic of passive use value is that there is no observable related behavior from which to judge people’s willingness to make tradeoffs for the benefit in question. The role of the stated preference survey, explicitly, is to create a situation in which we can observe this willingness. Hausman’s argument is that because stated preference surveys cannot reliably elicit information on preferences, then an expert panel is preferred. The problem is that no expert can claim to know the preferences of a particular population without having first made at least some informal observations about people’s willingness to make tradeoffs. Although set in a different context, Smith’s (2010) comment on the prospect of relying on expert panels is worth noting: “This suggests that all we need to do is convene an expert panel that will somehow come to a consensus. I have lost count of how many conferences and conference sessions have been charged with this task and have failed to arrive at a consensus value.”

Surveys of experts, known as Delphi surveys, have been used in other disciplines for eliciting baseline estimates of population quantities of interest. Under this method, a group of experts reply to questionnaires and then receive feedback in the form of a statistical representation of the group response, after which the process repeats itself. The goal is to reduce the range of responses and arrive at something closer to expert consensus. The use of Delphi surveys may offer an opportunity for an increased role for expert opinion, but the only sure way to know if Delphi surveys provide accurate estimates of value is to also survey the population for whose value the experts endeavor to estimate. Given sufficient replications of accurate expert estimates, it may turn out to be a useful and potentially much cheaper valuation method. But here is the problem: Hausman argues that estimates elicited from the public are unreliable to begin with, which means it is impossible to ever discern if the Delphi method provides accurate value estimates or not. In the absence of such surveys of the public, as argued already, the so-called experts can have no baseline upon which to express their initial estimate of value, and if none of them have such information in hand, then the iterative

process of the Delphi method will simply result in increased precision where there is no accuracy, ultimately yielding a false “consensus” of value.

Thus, the actual choice of passive use value estimates we are confronted with is not, as Hausman would have us believe, between those of a wholly unreliable stated preference survey and those of a reliable expert measure of passive use value. Instead, it is between the values advocated by a statistically uninformed, highly selective and potentially biased “expert” panel, and those implied by a stated preference survey that is possibly flawed, but nevertheless broadly and systematically informed (in the sense of actually being based on direct input from a large representative sample drawn from the population of interest).

If Hausman is instead arguing only that the possible existence of passive use values should be acknowledged by these experts but that they should go no further in quantifying the magnitudes of these values, then his advice is even less useful. Mere acknowledgement of passive use values cannot be incorporated into a decision process in any systematic and meaningful way. Given the mandate in the United States for formal benefit-cost analyses for significant policies or regulations (Smith 1984), failure to monetize the whole category of passive use benefits creates the risk that the value of these benefits will be defaulted to zero in arriving at the bottom line of a net benefits calculation. When expert opinions about values are substituted in lieu of stated preference evidence, it is inevitable that the process of high-stakes policy-making will produce challenges to the basis for their expertise about societal values. Evidence-based policy-making requires just that: evidence. The comparison of all the costs of an action to all the benefits of an action is the decision criterion that separates economics from other disciplines. We acknowledge that using contingent-valuation surveys to estimate the value of some things, such as certain obscure ecosystem services, presents a challenge; giving up on the proper measurement and incorporation of some (potentially sizeable) benefits, however, means giving up on proper BCA. BCA should not necessarily be the only criterion used in decision-making, but it is the economic one. And although it is always necessary to determine how much weight BCA actually ends up having in any decision process, Hausman has no greater authority than anyone else in determining whether passive use values should or should not be formally incorporated into the decision process.

*Altruism and BCA.*

Hausman also argues that values deriving from altruism should not be incorporated into BCA, and that it is difficult to identify and exclude altruistic values from stated preference studies. There are two main classes of altruism, non-paternalistic and paternalistic (Jacobsson et al. 2007), with the former generally causing more theoretical difficulty in CVM. Non-paternalistic altruism is when one cares about another’s general well-being (i.e., utility level), whereas paternalistic altruism is when one cares about another’s level of a particular component of utility. The risk with non-paternalistic altruism is that a single person’s utility is captured more than once in the value estimate (because it appears in the utility functions of others). Bergstrom (1982) showed that the rule for the Pareto efficient level of a public good (i.e., the Samuelson rule) is the same in the presence or absence of non-paternalistic altruism. This implies that we require

purely “selfish” values—excluding altruistic values—for determining optimal outcomes. However, one of the assumptions made by Bergstrom was that the considered change be small and that the economy already be near the Pareto efficient level of the public good. Flores (2002) showed that for discrete (large) changes in the public good, such as are often the subject of CV studies, the selfish benefit-cost test is sufficient but not necessary for making an efficient decision. That is, it is possible to fail to accept a good proposal (from a BCA point of view) using only selfish values. Bergstrom (2006) agrees with this conclusion.

Non-paternalistic altruism causes difficulty because in its presence, the benefits of a project actually depend on the distribution of the costs of the project (Flores 2002). To see why, consider a thought experiment involving two projects with identical expected outcomes and total costs of implementation, the second of which places a heavier cost burden on someone for whom the respondent feels non-paternalistic altruism. For this respondent, the benefits of the first project are larger than the benefits of the second project. Curious researchers can investigate ways to detect altruism in WTP or choice question responses and, if possible, to develop value estimates that both include and exclude the altruistic component for comparison. Such researchers can also investigate the way in which the distribution of costs of a proposed project influences value estimates.

*Unstable and Inconsistent Preferences*

A recurring theme throughout Hausman’s (2012) paper is the assertion of unstable preferences as an indictment of CVM methods. For example, on page 53 he writes:

> . . . the primary argument that is relevant for thinking about contingent valuation methods as awhole is that the answers from such studies are unstable and inconsistent, invented for the moment of the survey, and cannot be treated as preferences in the sense that economist understand that term.

Although the papers in the Journal of Economic Perspectives symposium focus mostly on the validity of the CVM (e.g., hypothetical bias and scope), the assertion that CVM responses are “unstable and inconsistent” is a statement about their reliability, not their validity. Hausman does not properly support his assertion about the unreliability of CVM with references to the literature.

Validity is the extent to which a valuation method generates a measure that is unbiased, that is, an estimate centered around the true value, if it were known. Validity is difficult to demonstrate when valuing nonmarket goods and services because, by their nature, their “true” value is unknown. A valid method for estimating these values is thus one that attempts to provide an unbiased estimate around an unknown and unobservable quantity.

Reliability is the extent to which a valuation method consistently generates the same measure. Reliability tests focus on the within and across study variation in estimates rather than the ability of studies to produce unbiased estimates of value. The lower the variability in estimates, the more consistent and the less influenced by researcher decisions are the estimates. High variability allows seemingly innocuous

decisions by practitioners to significantly influence results in one direction or another, so unreliable results cannot be defended against the criticism that the researcher may be determining the result.

Most CVM research finds that WTP values have some degree of stability over time (e.g., Brouwer 2006). Test-retest (temporal) reliability involves conducting more than one CVM survey with time between surveys using the same or a different sample of respondents. If the magnitude of WTP is consistent across time, then WTP is considered temporally reliable (e.g., Carson et al. 1997). However, a difference in WTP over time does not necessarily indicate unreliable results. If WTP changes over time in response to changing factors that would be expected to affect WTP, such as bubbles in housing markets or a significant recession, then the researcher may still conclude that the results are reliable (e.g., Whitehead and Hoban 1999). Overall, it is important that CVM studies demonstrate some degree of both validity and reliability. We are also not convinced that unstable or constructed preferences are unique to CVM. The formation of preferences has to begin somewhere. The formation of preferences on the spot in a survey does not necessarily invalidate them for those who have no prior preferences or experience. When confronted with unfamiliar but interesting products in real markets, people also sometimes make impulsive, spontaneous, or ill-advised purchases that they may subsequently regret. Perhaps the preferences elicited by CVM are in their “infancy” stage, but this can be true for market goods and services as well (new products are introduced all the time). It is also probably safe to say that (almost) everyone has well-formed preferences for money, so one can at least make a “no-purchase” decision based on well-defined preferences for “all other goods” relative to the proposed public good with unfamiliar qualities (expressed by a vote against the provision of that good in a consequential referendum).

## Conclusions

Hausman (2012) makes an effort to debunk the CVM once and for all, thereby removing it from the research agenda and from use in litigation and public policy. A compelling counterargument to Hausman’s claims is to note that Hausman takes the results of many of his selected studies at face value. While compelling, a similar claim can be made about many of the reflections made in this paper; some of the works cited are selected to provide a specific, non-critical, face value, counterargument to points made by Hausman, while others are examined more critically to provide the curious reader with an opportunity to examine for themselves some of the unanswered questions that remain in the ongoing CVM debate. While critical examination of every paper cited herein, or in the broader literature for that matter, is a worthy goal, the vastness of the literature, and space limitations, prevent a comprehensive critical analysis here.

Nevertheless, in direct response to Hausman’s selective interpretation of the literature, we believe that the overwhelming amount of evidence shows: (1) the existence (or nonexistence) of hypothetical bias continues to raise important research questions about the incentives guiding survey responses and preference revelation in real as well as hypothetical settings, and contingent valuation can help answer these questions; (2) the WTP-WTA gap debate is far from settled and raises important

research questions about the future design and use of benefit cost analyses in which contingent valuation will undoubtedly play a part; and (3) CVM studies do, in fact, tend to pass a scope test and there is little support for the argument that the adding up test is the definitive test of CVM validity.

We are in complete agreement with Carson (2012), who concludes “the time has come to move beyond endless debates that seek to discredit contingent valuation and to focus instead on making it better.” Proponents of the CVM and other stated preference methods should continue to make cautious claims about the accuracy of CVM results. As the stated preference literature has expanded from simple CVM questions into the realm of choice experiments, contingent behavior and experimental economics, many of the remaining questions about CVM continue. Choice experiments should be given closer scrutiny to determine if they are an improvement over, or simply an alternative to CVM, and if they have better incentive properties than well-designed CVM questions.

Throughout this reflection, we have highlighted a multitude of unanswered questions that remain for the curious, some of which include the following: Do incentive properties of questions affect the WTP/WTA gap? How do we, as researchers, determine whether respondents face the same set of incentives we believe they face? Is consequentiality measurable, and if so, how do we measure it? In the estimation and calculation of WTP, are we controlling for the behavioral influences and deviations from assumptions about such behavior that we think we are controlling for? In the context of natural resource damage assessment, CVM-based damage assessments are useful as starting points for damage awards. In benefit-cost analysis, CVM estimates are useful as baselines or bounds around which sensitivity analysis should be conducted. Our hope is that the economics profession does not blindly accept the flawed “dubious to hopeless” pronouncement.

## Acknowledgments

The authors wish to thank Trudy Cameron, Richard Carson, John Loomis, Ted McConnell, Kerry Smith, and an anonymous reviewer for their suggestions, which greatly improved this manuscript.

## References


Amiran, E.Y., and D.A. Hagen. 2010. The Scope Trials: Variation in Sensitivity to Scope and WTP with Directionally Bounded Utility Functions. Journal of Environmental Economics and Management 59 (3): 293–301.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." Federal Register 58, no. 10 (15 January 2008).

Bergstrom, T.C. 1982. When is a Man's LifeWorth More than his Human Capital? In: The Value of Life and Safety: Proceedings from a Conference Held by the Geneva Association, ed. M.W. Jones-Lee, 3–25. Amsterdam: North Holland Press.
———. 2006. Benefit-cost in a Benevolent Society. American Economic Review 96 (1): 339–351.

Blumenschein, K., G.C. Blomquist, M. Johannesson, N. Horn, and P. Freeman. 2008. Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment. The Economic Journal 118 (525): 114–137.

Bulte, E., S. Gerking, J.A. List, and A. de Zeeuw. 2005. The Effect of Varying the Causes of Environmental Problems on Stated WTP Values: Evidence from a Field Study. Journal of Environmental Economics & Management 49 (2): 330–342.

Brouwer, R. 2006. Do Stated Preference Methods Stand the Test of Time? A Test of the Stability of Contingent Values and Models for Health Risks when Facing an Extreme Event. Ecological Economics 60: 399–406.

Cameron, T.A., and J.R. DeShazo. 2013. Demand for Health Risk Reductions. Journal of Environmental Economics and Management 65 (1): 87–109.

Carlsson, F., M. Kataria, A. Krupnick, E. Lampi, A ° . Lo¨fgren, P. Qin, and T. Sterner. 2013. The Truth, the Whole Truth, and Nothing but the Truth – A Multiple Country Test of an Oath Script. Journal of Economic Behavior & Organization 89: 105–121.

Carson, R.T. 2011. Contingent Valuation: A Comprehensive Bibliography and History. Cheltenham, UK: Edward Elgar.

———. 2012. Contingent Valuation: A Practical Alternative when Prices Aren't Available. Journal of Economic Perspectives 26 (4): 27–42.

Carson, R.T., and T. Groves. 2007. Incentive and Informational Properties of Preference Questions. Environmental and Resource Economics 37 (1): 181–210.

———. 2011. Incentive and Information Properties of Preference Questions: Commentary and Extensions. In International Handbook of Non-Market Environmental Valuation, ed. J. Bennett, 300–321. Northampton,MA: Edward Elgar.

Carson, R.T., W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, et al. 1997. Temporal Reliability of Estimates from Contingent Valuation. Land Economics 73 (2): 151–163.

Champ, P.A., and R.C. Bishop. 2001. Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias. Environmental and Resource Economics 19 (4): 383–402.

Cummings, R.G., and L.O. Taylor. 1999. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. American Economic Review 89 (3): 649–665.

Desvousges,W., K. Mathews, and K. Train. 2012. Adequate Responsiveness to Scope in Contingent Valuation. Ecological Economics 84 (December): 121–128.

Diamond, P. 1996. Testing the Internal Consistency of Contingent Valuation Surveys. Journal of Environmental Economics and Management 30 (3): 337–347.

Diamond, P.A., and J.A. Hausman. 1994. Contingent Valuation: Is Some Number Better than No Number? Journal of Economic Perspectives 8 (4): 45–64.

Flores, N.E. 2002. Non-paternalistic Altruism and Welfare Economics. Journal of Public Economics 83 (2): 293–305.

Grijalva, T.C., R.P. Berrens, A.K. Bohara, and W.D. Shaw. 2002. Testing the Validity of Contingent Behavior Trip Responses. American Journal of Agricultural Economics 84 (2): 401–414.

Hanemann, W.M. 1991. Willingness to Pay and Willingness to Accept: How Much Can They Differ? American Economic Review 81 (3): 635–647.

———. 1994. Valuing the Environment through Contingent Valuation. Journal of Economic Perspectives 8 (4): 19–43.

Hausman, J.A. 1993. Contingent Valuation: A Critical Assessment. Amsterdam: North Holland.

———. 2012. Contingent Valuation: From Dubious to Hopeless. Journal of Economic Perspectives 26 (4): 43–56.

Herriges, J., C.L. Kling, C.C. Liu, and J. Tobias. 2010. What are the Consequences of Consequentiality? Journal of Environmental Economics & Management 59 (1): 67–81.

Hillygus, D.S. 2011. The Evolution of Election Polling in the United States. Public Opinion Quarterly 75 (5): 962–981.

Hsiao, C., B. Sun, and V.G. Morwitz. 2002. The Role of Stated Intentions in New Product Purchase Forecasting. In Advances in Econometrics, Vol. 16, ed. T. Fomby, R. Carter Hill, I. Jeliazkov, J.C. Escanciano, and E. Hillebrand, 11–28. Bingley, UK: Emerald Publishing Group Limited.

Jacobsson, F., M. Johannesson, and L. Borgquist. 2007. Is Altruism Paternalistic? The Economic Journal 117 (520): 761–781.

Jacquemet, N., R.V. Joule, S. Luchini, and J.F. Shogren. 2013. Preference Elicitation under Oath. Journal of Environmental Economics and Management 65 (1): 110–132.

Jamieson, L.F., and F.M. Bass. 1989. Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: AComparison of Models and Methods. Journal of Marketing Research 26 (3): 336–345.

Johnston, R.J. 2006. Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum. Journal of Environmental Economics and Management 52 (1): 469–481.

Kling, C.L., D.J. Phaneuf, and J. Zhao. 2012. From Exxon to BP: Has Some Number Become Better than No Number? Journal of Economic Perspectives 26 (4): 3–26.

Knetsch, J. 2010. Values of Gains and Losses: Reference States and Choice of Measure. Environmental and Resource Economics 46 (2): 179–188.

Krutilla, J.V. 1967. Conservation Reconsidered. American Economic Review 57 (4): 777–786.

Landry, C., and J.A. List. 2007. Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field. American Journal of Agricultural Economics 89 (2): 420–429.

List, J.A. 2001. Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auction Experiments. American Economic Review 91 (5): 1498–1507.

List, J.A., and C.A. Gallet. 2001. What Experimental Protocol Influences Disparities between Actual and Hypothetical State Values? Environmental and Resource Economics 20 (3): 241–254.

Little, J., and R. Berrens. 2004. Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis. Economics Bulletin 3 (6): 1–13.

Loomis, J.B. 2011. What's to Know about Hypothetical Bias in Stated Preference Valuation Studies. Journal of Economic Surveys 25 (2): 363–370.

McConnell, K.E., and J.K. Horowitz, 2003. Willingness to Accept, Willingness to Pay and the Income Effect. Journal of Economic Behavior and Organization 51 (4): 537–545.

Mitchell, R.C., and R.T. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method.Washington DC: Resources for the Future.

Morrison, G.C. 1998. Understanding the Disparity between WTP and WTA: Endowment Effect, Substitutability, or Imprecise Preferences? Economics Letters 59 (2): 189–194.

Morwitz, V.G., J.H. Steckel, and A. Gupta. 2007. When do Purchase Decisions Predict Sales? International Journal of Forecasting 23 (3): 347–364.

Murphy, J.J., P.G. Allen, T.H. Stevens, and D.Weatherhead. 2005. A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. Environmental and Resource Economics 30 (3): 313–325.

Ojea, E., and M.L. Loureiro. 2011. Identifying the Scope Effect on a Meta-analysis of Biodiversity Valuation Studies. Resource and Energy Economics 33 (3): 706–724.

Plott, C.R., and K. Zeiler. 2005. The Willingness to Pay–Willingness to Accept Gap, the Endowment Effect, Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. American Economic Review 95 (3): 530–545.

Portney, P.R. 1994. The Contingent Valuation Debate: Why Economists Should Care. Journal of Economic Perspectives 8 (4): 3–17.

Poe, G.L., and C.A. Vossler. 2011. Consequentiality and Contingent Values: An Emerging Paradigm. In The International Handbook on Non-Market Environmental Valuation, ed. J. Bennett, 122–141. Northampton, MA: Edward Elgar.

Randall, A., B. Ives, and C. Eastman. 1974. Bidding Games for Valuation of Aesthetic Environmental Improvements. Journal of Environmental Economics and Management 1 (2): 132–149.

Richardson, L., and J. Loomis. 2009. The Total Economic Value of Threatened, Endangered and Rare Species: An Updated Meta-Analysis. Ecological Economics 68 (5): 1535–1548.

Rollins, K., and A. Lyke. 1998. The Case for Diminishing Marginal Existence Values. Journal of Environmental Economics and Management 36 (3): 324–344.

Shogren, J.F., and D.J. Hayes. 1997. Resolving Differences in Willingness to Pay and Willingness to Accept: Reply. American Economic Review 87 (1): 241–244.

Shogren, J.F., S.Y. Shin, D.J. Hayes, and J.B. Kliebenstein. 1994. Resolving Differences
in Willingness to Pay and Willingness to Accept. American Economic Review 84 (1): 255–270.

Smith, V.K. 1984. Environmental Policy under Reagan's Executive Order: The Role of Benefit-cost Analysis. Chapel Hill: University of North Carolina Press.

———. 2010. Reflections – Legacies, Incentives, and Advice. Review of Environmental Economics and Policy 4 (2): 309–324.

Smith, V.K., and L.L. Osborne. 1996. Do Contingent Valuation Estimates Pass a "Scope" Test? A Meta-Analysis. Journal of Environmental Economics and Management 31 (3): 287–301.

Sugden, R. 2005. Coping with Preference Anomalies in Cost–benefit Analysis: A Market-simulation Approach. Environmental and Resource Economics 32 (1): 129–160.

Sun, B., and V.G. Morwitz. 2010. Stated Intentions and Purchase Behavior: A Unified Model. International Journal of Research in Marketing 27 (4): 356–366.

Vossler, C., and M.F. Evans. 2009. Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality. Journal of Environmental Economics and Management 58 (3): 338–345.

Vossler, C.A., M. Doyon, and D. Rondeau. 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economics Journal: Microeconomics 4 (4): 145–171.

Vossler, C.A., and S.B. Watson. 2013. Understanding the Consequences of Consequentiality: Testing the Validity of Stated Preferences in the Field. Journal of Economic Behavior and Organization 86 (February): 137–147.

Weaver, R., and D. Prelec. 2013. Creating Truth-telling Incentives with the Bayesian Truth Serum. Journal of Marketing Research 50 (3): 289–302.

Whitehead, J.C. 2005. Environmental Risk and Averting Behavior: Predictive Validity of Jointly Estimated Revealed and Stated Behavior Data. Environmental and Resource Economics 32 (3): 301–316.

Whitehead, J.C., T.C. Haab, and J-C. Huang. 1998. Part-whole Bias in Contingent Valuation: Will Scope Effects be Detected with Inexpensive Survey Methods?
Whitehead, J.C., T.C. Haab, and J-C. Huang., eds. 2011. Preference Data for Environmental Valuation: Combining Revealed and Stated Approaches. New York: Routledge.

Whitehead, J.C., and T.J. Hoban. 1999. Testing for Temporal Teliability in Contingent Valuation with Time for Changes in Factors Affecting Demand. Land Economics 75 (3): 453–465.